UNCLASSIFIED

AD **296 428**

Reproduced
by the

ARMED SERVICES TECHNICAL INFORMATION AGENCY
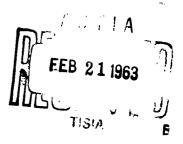ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA

UNCLASSIFIED

63-2-4

TECHNICAL NOTE NO. 6

under Contract
AF49(638)-1128

"Inductive Methods in Semantic
Analysis"

15 January 1963

THOMPSON RAMO WOOLDRIDGE INC.

RW DIVISION

CANOGA PARK, CALIFORNIA

$2.60

RESEARCH IN SEMANTIC STRUCTURE[1]

Paul L. Garvin

Thompson Ramo Wooldridge Inc.
RW Division

The greatest problem in the field of semantics today is our lack of reliable knowledge of semantic systems. In my opinion, this can be remedied only by the application of an empirical method to the processing of semantic data. It is possible to develop such a semantic research method by using the techniques of linguistic field work as a point of departure. I shall discuss my present efforts in this direction.

It is possible to envision a systematization of meaning not unlike the systematization of linguistic form which is done in descriptive linguistics. It is likewise not unreasonable to assume that some of the methods which have allowed us to systematize formal linguistic relations may be contributory to a systematization of semantic relations. The following linguistic considerations have bearing on such a systematization.

First of all, the basic assumption that there exists for each language a system of meanings comparable to the system of forms, allows a systematic attack upon the problem of meaning by linguistic methods.

Secondly, two consistent observations underlie the formula-
tion of linguistic techniques for the treatment of meaning:

(1) That, irrespective of theoretical controversies about the
"nature" of meaning, there are two kinds of observable and opera-
tionally tractable manifestations of linguistic meaning—translation
and paraphrase.

(2) That linguistic units with similar meanings will tend to
occur in environments characterized by certain specifiable
similarities.

The first observation allows the formulation of form-meaning
techniques[2] for semantic classification based on similarities in the
translation on paraphrase of the content-bearing elements of a
language. In a previous study[3] I have dealt extensively with the
use of translation as an operational representation of meaning.
The present study is concerned with paraphrase.

The second observation suggests the extension of linguistic
techniques of distributional analysis from problems of form to
problems of meaning.

The purpose of these techniques is to arrive at a semantic
classification of the content-bearing elements of a language which
is inductively inferred from the study of a sample and which there-
fore may be expected to reflect the intrinsic structure of the
language rather than some viewpoint external to it. To achieve
this aim, the conditions affecting the application of the techniques

---

[2] A form-meaning technique is an analytic technique drawing upon
the co-variance of form and meaning in language. It is here
opposed to a distributional technique.

[3] Paul L. Garvin, "A descriptive technique for the treatment of
meaning," Language 34.1-32 (1958).

must be controlled and appropriate comparison properties insuring their correct application must be specified.

An operation of paraphrasing can be defined as one in which an original statement or a part thereof is replaced by another statement or part statement. It will be considered right if the statement resulting from the paraphrase is in some relevant way comparable to the original statement.

In terms of the above, the following conditions have to be stipulated in order to specify a particular paraphrasing technique:

(1) the original statement;

(2) the portion of the statement (which may include all of it) that is to be paraphrased (the original form);

(3) the statement or part statement replacing the original form as a result of the paraphrase (the replacement form);

(4) the statement resulting from the paraphrase (the resultant statement);

(5) the particular property or properties in terms of which the resultant statement is compared to the original statement to ascertain whether the paraphrase is right (the comparison property).

A distributional technique can be defined as one based on the examination of the particular distributional conditions under which a given linguistic unit is found in a sample of the language. The following conditions thus have to be stipulated to specify a particular distributional technique:

(1) the type of linguistic units to be examined (the current forms);

(2) the units in the context of the current forms that are relevant to the purpose of the investigation—in our case, semantic research (the contextual definers).

As can be seen, a distributional technique requires the prior availability of the contextual definers in terms of which the current forms are to be described. These can be arrived at by the prior use of a form-meaning technique. Logically, the application of form-meaning techniques therefore precedes the application of distributional techniques. Likewise, the form-meaning techniques should be applied to linguistic units of a class well suited to serve as contextual definers for units of another class, to which then the distributional techniques can be applied.

The semantic research now in progress uses a form-meaning technique: a specially defined paraphrasing technique. This technique is applied to the predicative elements of a language, for two reasons:

(1) predicative elements can be assumed to be functionally independent clause members and hence to constitute a more suitable point of departure than the nominal elements which can be assumed to be function-dependent on them;

(2) once the predicative elements have been analyzed semantically, they can be expected to constitute a suitable set of contextual definers for applying a distributional technique to the semantic analysis of nominal elements, precisely because of the function dependence mentioned above.

The technique, called predication-typing, has been developed in connection with the analysis of the content of technical and journalistic writing, since this type of writing is less likely to be affected by the poetic function than, for instance, literary text, and hence can be expected to be semantically more straightforward.

On the basis of an impressionistic inspection of a fair-sized body of technical writing, the assumption was made that a large number of sentences in these documents could be described as presenting the functional relation of two terms, if this relatedness was conceived broadly enough (A is in some way a function of B). Note that the presence of both an "A" term and a "B" term are required in order for this generalized relation to be observed. In line with this assumption, it was found possible to reword a large majority of the sentences of the sample in a standardized form as shown above. As an example, the sentence found in our sample (see example (5) in Table 3):

"The value for fluorine comes
from the 1934 paper of the
Noddacks"

could be reworded in a most general way to read:

"The value for fluorine is in
some way a function of the
1934 paper by the Noddacks"

Any sentence lending itself to this type of rephrasing was termed a predication; sentences not lending themselves to it were termed non-predications.

It was found that the very generalized predication "A is in some way a function of B" could be further specified in terms of a limited number of relations which are intended to be semantic in the sense of being part of the cultural system of reference, rather than logical in the sense of being part of a formal system. These relations were called predication types. Each of these could be represented by a suitable replacement sentence type, to be used in rewording the running sentences of the documents. The predication types were selected intuitively in the course of a further inspection of text.

5

In the selection of the predication types, the aim of the technique, namely that of serving in the semantic analysis of technical documents, was kept in mind. In order to relegate as much of the non-technical content of the documents as possible to non-predications, the predication types and the corresponding replacement sentence types were formulated in such a way as to apply to relations between objects and concepts more readily than to actions performed by people. One of the by-products of the technique thus became a gross separation of technical from non-technical language.

A paraphrasing technique could thus be specified as follows: the replacement forms would be the predicative elements contained in the replacement sentence types; the original forms would be those forms in the original statement on the basis of which the particular replacement sentence types were chosen. Thus, to the sentence cited above, the predication type "indication of basic relation" could be applied. The replacement sentence for this predication type is "A is based on B"; applying it to the original statement we obtain:

> "The value of fluorine is
> based on the 1934 paper
> by the Noddacks"

The original form "comes from" has been replaced by the replacement form "is based on" in such a way as to satisfy the comparison property which for predication-typing is the approximate sameness of meaning of the original statement and the resultant statement. This comparison property is ascertained on the basis of naive native speakers' ex post facto judgments, as shown by responses to the question: "Does the resultant statement mean roughly the same as the original statement?" Informants made familiar with the purpose of the research were able to react to this test question with reasonable consistency. In the case of our example, the informant response was positive. This was held to satisfy the comparison property and to insure that the paraphrase is right.

The use of replacement forms makes predication-typing comparable to the substitution technique. It is worth noting the significant differences between the two techniques:

(1) In substitution as I practice it, [4] the frame is held constant and the number of permitted replacements is limited only by the requirement of right substitutability as expressed by the comparison property. In predication-typing, on the other hand, the number of permitted replacement forms is restricted and held constant, while the equivalent of the frame (that is, the portion of the statement not included in the replacement form) is modifiable within the limits of right rephrasing as expressed by the comparison property.

(2) In the substitution technique, the comparison property for right substitutability is grammatically defined, e.g., by dependences or clause functions. In predication-typing, the comparison property for right rephrasing is defined not grammatically, but semantically.

The application of predication-typing to the analysis of the semantic system of a language is based on the assumption that each predication type represents a semantic feature of the original form that has been rephrased by means of the replacement form belonging to that type. A given form of the language will then have as many semantic features as the number of predication types that have been used to rephrase it in the processing of a suitable sample of the language. On the basis of the semantic features obtained,

---

[4]Cf. Paul L. Garvin, "Syntactic units and operations, " Proceedings of the VIII International Congress of Linguists, pp. 630-1 (Oslo, 1958), and "A study of inductive method in syntax, " Word 18. 114-20 (1962).

the many forms of a language can then be arranged systematically. This arrangement will then represent the portion of the semantic system of the language corresponding to these forms.

The following example (see example (4) in Table 3) illustrates how replacement forms can serve to define the semantic features of a given original form.

Three predication types have been recorded for sentences containing the original form "start." A sample sentence for each predication type is analyzed below.

Rephrasing by first predication type.

Original statement:	Species when domesticated start to vary
Original form:	start
Replacement form:	be followed by
Resultant statement:	The domestication of species is followed by variation
Semantic feature inferred from operation:	succession

Rephrasing by second predication type.

Original statement:	Our own studies were started as part of a plan
Original form:	start
Replacement form:	stem from
Resultant statement:	Our own studies stemmed from a plan
Semantic feature inferred from operation:	origin

Rephrasing by third predication type.

Original statement:	Starting with the right naturally occurring amino acids, our group set out on the arduous task of copying nature

8

Original form:	start
Replacement form:	be based on
Resultant statement:	The arduous task of copying nature was based on the right naturally occurring amino acids
Semantic feature inferred from operation:	basic relation

The form "start" thus can be said to have, on the basis of the three types of rephrasing that have been recorded, three semantic features: succession, origin, and basic relation (see example (4) in Table 2).

The technique of predication-typing as illustrated above has by now been applied to a total of 40 documents, containing a total of approximately 4500 sentences. A total of 33 predication types (see Table 1) has so far been found sufficient.

The processing of the documents was conducted by attempting to rephrase every sentence of each document in line with the replacement sentence type of one of the 33 predication types. If a given sentence in a document did not lend itself to predication-typing it was considered a non-predication (symbolized by $\emptyset$) and left for later analysis by other means. The work of assigning predication types by rephrasing was carried out by myself and two assistants, one of them a nonlinguist. For reasons of practicality, the comparison property was tested by spot check only, rather than individually for each sentence.

The sample of 4500 sentences was clearly not sufficient to permit a detailed analysis of the semantic system of the technical language represented in it, since no single speech form occurred with sufficient frequency to allow inferences as to its detailed placement in a semantic structure. The sample was, however, sufficient to yield a number of preliminary results showing the bearing of the technique of predication-typing on certain presumed general characteristics of the semantic system of a language.

I should like to present these results. The analysis is summed up in Table 2. A portion of the corresponding research record is reproduced for purposes of verification in Table 3.

The assumption that links predication-typing to semantic structure has been stated further above: that each predication type represents a semantic feature of the original form that has served as a basis for the selection of that predication type. This assumption allows us to utilize the results of predication-typing in the analysis of two semantic dimensions.

Firstly, taking the replacement forms as a constant, we can investigate what original forms have served as a basis for their use. It can then be asserted that all of these original forms are to some degree synonymous with respect to the semantic feature expressed by the predication type represented by the replacement form. The set of forms synonymous with respect to the semantic feature expressed by predication type F, "statement of succession," is shown in the column of forms in Table 2.

Secondly, taking the original forms as a constant, we can investigate the predication types that have been selected on the basis of each original form. It can then be asserted that the semantic features expressed by these predication types constitute the semantic spectrum of the original form in question. The semantic spectra of the forms mentioned further above are shown by the rows of symbols adjacent to each form in Table 2.

The number of semantic features represented in the spectrum of a given form can then serve as a rough indication of the extent of the range of meaning of that form. Compare on that basis the ranges of meaning of the forms (3) result and (6) find shown in Table 2.

A more accurate measure of both synonymy and semantic spectrum can be envisioned if one considers not only what forms have been rephrased and what predication types have been used in the rephrasing, but also the relative frequency of occurrence of

10

these. It will then be possible to state not only whether a given predication type has occurred as part of the semantic spectrum of a form, but also—based on the relative frequency of occurrence—show the weight of this predication type within the spectrum. In regard to synonymy, it may then become possible to develop a measure for the degree of synonymy of two forms with respect to a predication type, based on the relative weights of this type in the semantic spectra of the two forms. Ultimately, we may envision an overall measure of synonymy based on the relative weights of all shared predication-type derived semantic features, and this overall measure can then be thought of as the basis for a precise representation of this aspect of semantic structure.

To be reliable, the simplest of the above measures will require a sample far in excess of the one processed to date. Only then will it be possible and worth while to work out the statistical details of these measures, and to perform some calculations.

The existing sample allows no more than a superficial impression of what might be the degree of synonymy of some pairs of forms selected on the basis of non-minimal frequency of occurrence in the sample. Compare, for instance, the degree of synonymy with respect to predication type O, semantic feature "indication of origin" ("A stems from B") of the pair come--result with that of the pair come--find.

In addition to the above-discussed suggestive but far from conclusive results, one definitive result emerged from the research conducted so far. Contrary to expectation, it turned out that the original forms—those on the basis of which the replacement sentence types were chosen—were not limited to predicative words and phrases, in spite of the fact that all of the actual replacement forms were predicative in character (see Table 1). On the contrary, the original forms include words of all classes, among them conjunctions and prepositions. This is due to the fact that sometimes no rephrasing based on one of the 33 predication types could be based on the

predicative elements contained in a particular original statement.
In many of these examples it seems counterintuitive that no predication in our sense should be present. If we assume that a predication is present covertly, it may be revealed when the rephrasing is based on a non-predicative original form. An example is the original statement:

"Once freezing is initiated, the drops grow ..."

If the rephrasing is based on the predicates of the two clauses composing this sentence, two non-predications will result, since every predication by definition requires both an "A" term and a "B" term, and neither clause can be rephrased on the basis of its predicate in such a way that it will satisfy this condition. A predication can, however, be achieved if the rephrasing is based on the non-predicative element "once"; we then obtain:

"The initiation of freezing is followed by the
growth of the drops..."

which is clearly a type F predication.

The variety of word classes to which original forms may belong appears from an inspection of the column of forms in Section 3. of the handout.

Two tentative conclusions can be drawn from this. One is that the semantic classification revealed by predication-typing does not coincide with the formal grammatical division into word classes. The second is that function words — such as conjunctions and prepositions — when viewed in the light of predication-typing turn out to have significant elements of meaning in common with content words — such as certain verbs and nouns. An example of this is the semantic spectra of the function word (20) from and the content word (3) result shown on the handout.

Table 1. List of Predication Types

Symbol	Name of Predication Type	Replacement Form	
A	statement of accompanying circumstance	A is accompanied by	B
Ac	statement of acquisition	A is acquired by	B
B	indication of basic relation	A is based on	B
C	statement of causation	A is due to	B
Cm	statement of computation	A is computed for	B
Cn	statement of concern	A concerns	B
Cp	statement of comparability	A is comparable to	B
Cs	indication of constituency	A constitutes	B
D	descriptive statement	A has a property of	B
Df	definition	A is defined as	B
Dm	demonstration	A is shown by means of	B
E	description of equipment	A is provided with	B
Ef	indication of effect	A affects	B
Ex	explanatory statement	A is accounted for by	B
F	statement of succession	A is followed by	B
I	indication of insertion	A is placed into	B
L	indication of location	A is found in	B
M	indication of measurement	A is measured by means of	B
Ma	description of movement away	A moves away from	B
O	indication of origin	A stems from	B
P	indication of possibility	A allows for	B
Pf	description of performance	A is accomplished by means of	B
R	statement of established result	A has been established as	B
Re	indication of replacement	A is replaced by	B
Rm	description of result of motion	A arrives at	B
Rp	statement of represented condition	A is presented as	B

Table 1. List of Predication Types (Continued)

Symbol	Name of Predication Type	Replacement Form	
Rq	indication of requirement	A requires	B
S	assertion of adequacy	A satisfies the conditions for	B
Su	indication of superiority	A exceeds	B
T	statement of passage	A passes through	B
U	description of use	A is used for	B
V	indication of co-variance	A varies with	B
Va	statement of value	A has the value of	B

14

Table 2. Preliminary Analysis

Predication Type F: A is followed by B

Original forms assigned to predication type F	Other predication types to which form at left has been assigned
Number of examples in text	Number of examples in text for each

(1)	follow		S	C	O	Ex			
	7		1	1	1	1			

(2)	conclude	P
	2	1

(3)	result	O	C	Ex	Pf	R
	1	21	11	1	1	1

(4)	start	O	B
	2	2	1

(5)	come	O	Rm	Ma	Va	B	S	L
	2	12	3	2	2	1	1	1

(6)	find	L	Cs	R	O	D	A	Ac	Rp
	1	8	5	5	4	2	2	2	2
		Va	Dm	Rm	I	M	S	U	Ø
		1	1	1	1	1	1	1	1

(7)	reply
	6

(8)	end	O	R
	1	1	1

(9)	not long in forthcoming
	1

(10)	that (N of time)
	4

(11)	adverb of time
	2

Table 2. Preliminary Analysis (Continued)

Number of examples in text	Number of examples in text for each

(12)	when		A	C	O	B	Cs	P
	12		29	8	6	4	2	1
(13)	then		O	P	C			
	17		3	1	1			
(14)	duly							
	1							
(15)	eventually							
	1							
(16)	finally							
	1							
(17)	immediately							
	1							
(18)	quickly							
	1							
(19)	subsequently							
	2							
(20)	from		O	B	Dm	Ma	A	
	1		14	3	2	2	1	
(21)	until		C	Re				
	1		2	1				
(22)	upon		A					
	1		1					
(23)	within		L	Cp				
	1		1	1				
(24)	to		U	Rm	A	Pf	Rq	Cn	I
	2		4	1	1	1	1	1	1

Table 2. Preliminary Analysis (Continued)

Number of examples in text		Number of examples in text for each								
(25)	on	Cn	L	A	Rm					
	2	9	3	1	1					
(26)	ahead	Su								
	1	1								
(27)	just	Cp								
	1	2								
(28)	again									
	2									
(29)	before									
	4									
(30)	after	P								
	30	2								
(31)	afterwards									
	3									
(32)	prior									
	1									
(33)	first	Rq								
	1									
(34)	later									
	7									
(35)	next									
	4									
(36)	as	A	Cp	C	Cs	B				
	1	8	8	3	2	1				
(37)	if	C	O	P	B	S	Rq	Ex	Pf	A
	1	51	22	9	4	2	2	1	1	1

Table 2. Preliminary Analysis (Continued)

Number of examples in text	Number of examples in text for each

		A	C	Cn	O	B	Ef
(38)	once	O					
	3	1					
(39)	since	C	O	P	B		
	3	54	3	2	1		
(40)	and	A	C	Cn	O	B	Ef
	4	4	2	1	1	1	1
(41)	but	A	C				
	2	4	1				
(42)	past gerund						
	1						
(43)	past participle						
	1						

Table 3. Research Record

Examples of Rephrasing

Original statement	Resultant statement

(1) follow

F: The emission is often closely followed by the emission of gamma rays   (same)

S: The radial distance . . . follows an A 1/3 distance law . . .   The radial distance . . . <u>satisfies the conditions for</u> an A 1/3 distance law

C: The prolonged effect that followed the injection of renin . . .   The prolonged effect that <u>was due to the injection of renin</u> . . .

O: It follows that the temperature is lower . . .   A lower temperature <u>stems from this</u> . . .

(2) conclude

F: Hays . . . concluded by placing in evidence . . .   This was followed by Hayes . . . placing in evidence . . .

P: From the arguments given here we conclude that . . . the likely places to look are the main sequence stars   The arguments given here <u>allow for</u> . . . the likely <u>places to</u> look to be in the main-sequence stars

(3) result

F: The ultimate result was that the ratio . . . was gauged   This was followed by the ratio . . . being gauged

O: A much greater polymorphism will result from recombination of the different genes   A much greater polymorphism will <u>stem from</u> recombination <u>of the different</u> genes

C: . . . the burdens South of the Equator rose as a result of the British test . . .   . . . the rising of the burdens South of the Equator <u>was due to</u> the British tests

Ex: The theory of the formation of cosmic rays as a result of supernovae . . . has been largely substantiated   The formation of cosmic rays <u>has been accounted for by</u> supernovae . . .

Table 3. Research Record (Continued)

Original statement	Resultant statement
Pf: The natural selection and evolution of organisms is a result of mutations	The natural selection and evolution of organisms is accomplished by means of mutations
R: The final result is a density even higher — unbelievable	The density has been established as even higher — unbelievable

(4) start

F: Species when domesticated start to vary	The domestication of species is followed by variation
O: Our own studies were started as part of a plan	Our own studies stemmed from a plan
B: Starting with the right naturally occurring amino acids, our group set out on the arduous task of copying nature	The arduous task of copying nature was based on the right naturally occurring amino acids

(5) come

F: ... then came the discovery of two stars	... this was followed by the the discovery of two stars
O: The additional power comes from the a-c pump	The additional power stems from the a-c pump
Rm: ... the particles come to earth ...	... the particles arrive at earth ...
Ma: ... power comes out of the reactor ...	... power moves away from the reactor ...
Va: ... the density comes out to more than 50,000 times the density of water ...	... the density has the value of more than 50,000 times the density of water ...
B: The value for fluorine comes from the 1934 paper of the Noddacks	The value for fluorine is based on the 1934 paper of the Noddacks
S: Both calculations ... come well within the limits of observational error ...	Both calculations ... satisfy the conditions for observational error ...

Table 3. Research Record (Continued)

Original statement	Resultant statement

L: The worst answer ... comes in the form of the Soviet index of industrial production

The worst answer ... <u>is found</u> in the Soviet index of <u>indus-trial</u> production

(6) find

F: ... his contributions ... found ready application

... his contributions ... <u>were</u> followed by ready applic<u>ation</u>

L: Such stones are found in strangely limited regions

(same)

Cs: Good examples ... are to be found among ornamental plants

Ornamental plants <u>constitute</u> good examples ...

R: The deductive method ... was found to have practical as well as esthetic value

The deductive method ... <u>has been established as having practical as well as</u> esthetic value

O: ... we can find in it some valuable insights ...

... some valuable insights <u>stem from</u> it ...

D: ... they are found to be identical ...

... they have the property of being identical ...

A: ... similar variations will be found in the brightness of this system

... the brightness of this system <u>will be accompanied by</u> similar va<u>riations</u>

Ac: ... he can perhaps find a substitute

... a substitute can perhaps <u>be acquired by</u> him

Rp: ... a man ... would find a trip through a crystal a tedious expedition

... a trip through a crystal <u>is presented as</u> a tedious expedition

Va: ... we find $\theta \sim 7$ minutes ...

... θ <u>has the value of</u> 7 minutes

Dm: Subtracting the average depth ... we find that the sediments range ... in thickness

The sediments' ranging ... in thickness <u>is shown by means of subtracting the</u> average depth ...

Table 3. Research Record (Continued)

Original statement	Resultant statement
Rm: arithmetic promptly found applicability ...	... arithmetic promptly arrived at applicability ...
I: ... Scopes found himself engaged in a discussion ...	... Scopes was placed into a discussion ...